

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2002, 23:33:11 ; Search time 1527.5 seconds  
(without alignments)  
16168.980 Million cell updates/sec

Title: US-10-025-514-15  
Perfect score: 1525  
Sequence: 1 tctagaccatggaagaccct.....ccagtaaggcctagtgcac 1525

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_othr:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429	28.1	2478	11 AF130068	AF130068 Homo sapi
2	427.4	28.0	2571	11 AF113676	AF113676 Homo sapi
3	331.2	21.7	1392	11 AK002537	AK002537 Mus muscu
4	318.4	20.9	1296	11 AK004999	AK004999 Mus muscu
5	309.8	20.3	887	11 BQ643710	BQ643710 AGENCOURT
6	305.4	20.0	907	14 BQ648909	BQ648909 AGENCOURT

7	286.8	18.8	895	14	BQ653587	AGENCOURT
8	285.2	18.7	982	14	BQ646142	AGENCOURT
9	284.2	18.6	985	14	BM924019	AGENCOURT
10	283.8	18.6	924	14	BQ958958	AGENCOURT
11	274.8	18.0	1194	14	BM924813	AGENCOURT
12	273.4	17.9	907	14	BQ650189	AGENCOURT
13	270.4	17.7	959	14	BQ648524	AGENCOURT
14	268.6	17.6	891	14	BQ646948	AGENCOURT
15	268	17.6	796	12	BG567260	AGENCOURT
16	266.4	17.5	833	14	BQ007663	AGENCOURT
17	264.8	17.4	931	14	BQ064738	AGENCOURT
18	263.6	17.3	687	10	AV649258	AV649258
19	262	17.2	686	12	BG569467	AGENCOURT
20	261	17.1	917	14	BQ850502	AGENCOURT
21	260.2	17.1	916	14	BQ644130	AGENCOURT
22	260.2	17.1	977	14	BQ654015	AGENCOURT
23	259.6	17.0	740	10	AV653778	AV653778
24	259.6	17.0	944	14	BQ645533	AGENCOURT
25	259.6	17.0	1012	14	BQ064473	AGENCOURT
26	256	16.8	869	13	BI219058	AGENCOURT
27	253	16.5	654	12	BG563537	AGENCOURT
28	251.4	16.5	884	12	BG36548	AGENCOURT
29	251.2	16.5	917	14	BQ650698	AGENCOURT
30	250.6	16.4	746	12	BG566873	AGENCOURT
31	250.2	16.4	745	9	AI114643	AI114643
32	249.8	16.4	1044	9	AI528087	AI528087
33	249.6	16.4	813	13	BI759000	BI759000
34	249.4	16.4	764	12	BF384318	BF384318
35	249	16.3	731	10	AV649144	AV649144
36	249	16.3	813	14	BQ182052	UI-H-EUO-
37	246	16.1	804	12	BG618951	BG618951
38	244.4	16.0	780	9	AI110706	HAQ145 HU
39	244.2	16.0	992	13	BI246204	BI246204
40	243.8	16.0	752	12	BG618001	BG618001
41	243.2	15.9	674	14	BM724546	UI-E-EQI-
42	243.2	15.9	759	12	BG533293	BG533293
43	243	15.9	749	9	AI526727	AI526727
44	242	15.9	629	12	BG545257	BG545257
45	241.2	15.8	780	12	BG617594	BG617594

## ALIGNMENTS

RESULT 1  
AF130068  
LOCUS AF130068 2478 bp mRNA linear HTC 08-MAY-2001  
DEFINITION Homo sapiens clone FLB8226 PRO2209 mRNA, complete cds.  
ACCESSION AF130068  
VERSION AF130068.1 GI:11493442  
KEYWORDS HTC  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2478)  
AUTHORS Zhang, C., Yu, Y., Zhang, S., Wei, H., Bi, J., Zhou, G., Dong, C., Zai, Y., Xu, W., Gao, F., Liu, M. and He, F.  
TITLE Functional prediction of the coding sequences of 75 new genes deduced by analysis of cDNA clones from human fetal liver  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2478)  
AUTHORS Zhang, C., Yu, Y., Zhang, S., Wei, H., Bi, J., Zhou, G., Dong, C., Zai, Y., Xu, W., Gao, F., Liu, M. and He, F.  
TITLE Direct Submission  
JOURNAL Submitted (23-FEB-1999) Department of Experimental Hematology, Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing, Beijing 100850, P. R. China  
FEATURES  
source Location/Qualifiers  
1..2478  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="FLB8226"









/gene="Sp11-4"  
/note="data source:MGD, source key:MGI:891968,  
evidence:ISS

putative  
serine protease inhibitor 1-4"

/codon\_start=1

/protein\_id="BAB23733.1"

/db\_xref="GI:12836612"

/db\_xref="MGD:MGI:891968"

/translation="MPSISWCLILLAGLCLVPSFLAEDVQETDTSOKDOSPASHEI  
ATNUGDAISLYRELHVHOSNTSNIFSPVSIATAPAMLSGSKGDTHQILLEGQNL  
TQTSFADSEHFKQLTNRPDSEIQLSTNGLNFVNDLKLVELEAKNIHQAEV  
FSVNFADSEHFKQLTNRPDSEIQLSTNGLNFVNDLKLVELEAKNIHQAEV  
KMOLEOTLAKELISKFLNMLRRRLAQIHIPRLSISGNVLETLMSPLGTRTFNSGA  
DLSGITENAPLKSQAVHKAVLTIDETGTEAANAIVLOGGFLSMPPLHFNRPFLFI  
IFEHSQSPFLVGRVDPDTHK"

BASE COUNT 339 a 370 c 300 g 287 t

ORIGIN

Query Match 20.9%; Score 318.4; DB 11; Length 1296;

Best Local Similarity 55.6%; Pred. No. 4.1e-72;

Matches 654; Conservative 0; Mismatches 516; Indels 6; Gaps 2;

QY 20 TCAAGCGACGCGCTCAAAAACCGACACCATCATCAGCACCAAGACCATCCGACTTT 79

Db 113 TCTGGCTGAGGATGTCAGGAGACAGACACCTCCAGAAAGGATCAGTCC---CCAGCCTC 169

QY 80 TAATAAATTAATCAAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAATTAGCTCA 139

Db 170 CCATGAGATTGTACAAACCTGGGAGACTTTGCCATCATCTATACCCGGAGCTGGTCCA 229

QY 140 TCAAGTAATTAATCAAAATTTTGTAGTCTTTTGTAGTCTTTTGTAGTCTTTTGTAGTCTTT 199

Db 230 TCAGTCCACACATTCACACATCTCTCTCCAGTGCAGCATTCGCCAGCCTTTGCTAT 289

QY 200 GTTGAGTTTGTAGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGGTTTAAACTTTAA 259

Db 290 GCTCTCCCTAGGAGCAAGGGTGACACTCACACGAGATCTCTAGAGGCGCTGCAGTTCAA 349

QY 260 TTTGACCGAATCCAGAACCCCAATTCACGAGGGTTTTCAGAGTTTGTTCAGAACTTT 319

Db 350 CCTCACACAAACATCGGAGGCTGACATCCACAAATTCCTTCCACACCTCTCCCAACCT 409

QY 320 GAATCAACTGATTTCAATTTCAATTAATCTAGTGAAGGTTTATTTTGTCTGAAAG 379

Db 410 CAACAGACGACAGTGCAGCTGCAGTGCAGCAGCAATGCTCTTTGTCAACAATGA 469

QY 380 TTTAAATTTGGTTGACAAATTCCTAGAGAGCTCAAGAACTATATCATAGTGGCTTT 439

Db 470 CCTGAGCTGGTGGAGAAGTTTCTGGAAGAGGCAAGAACCATTTATCAGGCAAGTCTT 529

QY 440 TACCGTTAATTTGGTGATCTGAGGAAGCTAAAGCAAAATTAATGATTTGTAGAA 499

Db 530 CTCTGTCAACTTTGAGAGTCAGAGAGGCGCAAGAAAGTGAATTAATGATTTGTGGGAA 589

QY 500 AGGCACCCAGGTTAGATCGTTGACCTAGTCTAAAGAAATTAGATCTGTATACCTCTCCG 559

Db 590 GGAACCCAGGAAGATAGTTGAGGCTGTGAAAGAACTAGAACAGACACAGTTTTCGT 649

QY 560 ACTAGTTAATTAATTTTCAAGGGTAAGTGGGACGCTCTCTTCGAGGTTTAAAGATAC 619

Db 650 CTGGCAAAATTAATTTCTTTAAAGGCAATGGAAGAGGCAATTCGATCTCTGAGAACAC 709

QY 620 TGAAGAGGAGATTTTCATGTGATCAAGTTACTACTGTCAAAGTTCCAAATGATCAAAAG 679

Db 710 TAAGAGCTGAGTTCACGTGGACGAGTCCACCAGGTGAGAGTCCCATGATGACCT 769

QY 680 ACTGGGTATGTTCAATTAATCAACATTTGCAAAAATTTAAGTTCTTGGGTCTTAAATGAA 739

Db 770 CTCGGGATGCTTTGAGCTGCACCATTTGCACAGCGCTCTCCAGCTGGGTCTGCTGATGA 829

QY 740 GTATTTAGGTAACGCTACTGCTATTTTCTTTTACCAGGAGGTTAAGCTTCAACATTT 799

Db 830 TTACGGGGCAAGCGCACCTGCTGCTTCTCTGCTGCCGATGATGGGAAGATCCAGCATCT 889

QY 800 AGAAGATGAGTTGACTCATGACATTAATTAATAATTTTATAGAAACGAGGATCGCTGATG 859

Db 890 GGAGCAAACTCAACAAGGAGCTCATCTTAAGTTCTGCTAAACAGGCGCAGAAGGTT 949

QY 860 CGTTCTCTGACCTGCCAAAGTAAAGTATACCGGTACTTACGACTTAAATCTGTTTT 919

Db 950 AGCCAGATCCATATCCAGAGCTGCTCATCTCTGGAATATTAATCTGGAGACACTCAT 1009

QY 920 AGCCAGTATGATTAACCAAAAGTTTTTCTAACGCTGCCGATTTGAGTGGTGTACTGA 979

Db 1010 GAGTCCACATGGGATCACCCGGGATCTTCAACAGTGGGGTGACCTTCCGGATCACAGA 1069

QY 980 AG---AAGCTCCATTAATAATAGTAAAGCTGTTCACAAAGCGCTTAACTATTGATGA 1036

Db 1070 GGAGATGCTCCCTCAAGCTCAGCCAGGCTGTGCATAAGGCGCTGCTGACCATCATGA 1129

QY 1037 AAGGGTACGAGCGCGCGCGCTATGTTCTGGAAGCTATTCATGAGCATTCACCTCCACC 1096

Db 1130 GACAGGAACAAAGCTGCAGCAGCTACAGTCTTACAAGGCGGTTTTTGTCTATGCCCCC 1189

QY 1097 AGAAGTAAATTAATAAAACCATTTCTGATGATCGAGCAGAACACTAAAAAGCCC 1156

Db 1190 TATCTGCACTTCAACGCCCTTCTCTTTCATTAATTTGAAGACACTCTCAGAGCCC 1249

QY 1157 ATTGTTTATGGTAAAGGTTGTCAACCCCACTCAGAA 1192

Db 1250 CCTTTTGTGGAAAGTGGTAGATCCACACATAA 1285

#### RESULT 5

BQ643710 887 bp mRNA linear EST 15-JUL-2002  
AGENCOURT\_8342217 NIH\_MGC\_100 Homo sapiens cdna clone IMAGE:6268225  
5' mRNA sequence.

#### ACCESSION

BQ643710

#### VERSION

BQ643710.1 GI:21767882

#### KEYWORDS

EST.

#### SOURCE

human.

#### ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

#### REFERENCE

1 (bases 1 to 887)

#### AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

#### TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

#### JOURNAL

Unpublished (1999)

#### COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgapb-1@mail.nih.gov](mailto:cgapb-1@mail.nih.gov)

Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LICM2439 row: n column: 02

High quality sequence stop: 731.

#### FEATURES

##### source

1..887

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6268225"

/tissue\_type="hepatocellular carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pOTB7; Site:1: XhoI; Site:2:  
ECORI; CDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)"



QY 577 TTTTCAAGGGTAAGTGGGAAGCTCTTTTCGAGGTTAAAGATACCTGAAGAGGAAGATTTTC 636  
 Db 375 TCTTTAAAGGCAATGGGAGAGACCCTTTGAAGTCAAGGACACCGAGGAAGAGACTTCC 434  
 QY 637 ATGTTGATCAAGTTACTACTGCTGAAGTTCCTCAATGATGAAGAAAGACTGGGTATGTTCAATA 696  
 Db 435 ACGTGGACCAAGTACCAACCGTGAAGGTGCGCTATGATGAAGCGTTTATAGGATGTTTAAACA 494  
 QY 697 TTTACATTTGCAAAAAATTAAGTTCTTGGGTCCTTATTAATGAAGTATTTAGGTAACGCTA 756  
 Db 495 TCCAGCATGTGAAGAAGCTGTCCAGCTGGGTGCTGTGATGAATACCTGGGCATGCCA 554  
 QY 757 CTGCTATTTTTTTTTTACAGACGAAGTAAAGCTTCAACATTTTAGAATGAGTTGACTC 816  
 Db 555 CCGCCATCTTCTCTGCTCATGAGGGAACCTACAGCACCTGGAAATGAATCACTCACCC 614  
 QY 817 ATGACATTTATTAATAATTTTAGAGACGAGGATCGTGTAGGCGTCTCTGCACTGC 876  
 Db 615 ACGATATCATCACCAAGTTCTCTGGAAATGAAGACAGAAAGGTCTGCCAGCTTACATTTAC 674  
 QY 877 CAAAGTTAAGTATCACCGGTACTTACGACTTAAATCTGTTTAAAGGCGAGTTAGGTATTA 936  
 Db 675 CCAAACTGTCCATCTACTGGAACCTATGATCTGAAGAGCGTCTGGGTCACTGGGCATCA 734  
 QY 937 CCAAAAGTTTTTCTAACGGTCCGATTTGAGTGGTGTACTGAAGAAAGCTCCATTTAAAT 996  
 Db 735 CTAAGTCTTCAGCAATGGGCTGACCTCTCCGGGTCAAGAGGAGGACCCCTGAAGC 794  
 QY 997 TGAGTAAAGCTGTTCAAAAGCGGTCTTAAGTATGATGAAGAGGTTACGAGGCCCGC 1056  
 Db 795 TCTCAGGAGG-CGTGATGAAGCTGTGCTGACCATCGACGAGAAAGGACTGAAGCTGCTG 853  
 QY 1057 GGGCTATGTTCTGGAAGCTATTCATGAGCATTCACACAGAAAGTT 1103  
 Db 854 GGGCCATTTGTTTGAAGGATACCATGCTAATCCCCCGAGGT 900

RESULT 7  
 B0653587  
 LOCUS  
 DEFINITION  
 AGENCOURT\_8349591 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6284550  
 5', mRNA sequence.

B0653587  
 VERSION  
 B0653587.1 GI:21777759  
 EST.  
 SOURCE  
 human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 895)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM2482 row: f column: 07  
 High quality sequence stop: 708.

FEATURES  
 source

1. 895  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6284550"  
 /clone\_lib="NIH\_MGC\_100"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2:

RESULT 8

ECoRT: cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 227 a 248 c 230 g 190 t  
 ORIGIN

Query Match 18.8%; Score 286.8; DB 14; Length 895;  
 Best Local Similarity 60.1%; Pred. No. 6.6e-64;  
 Matches 477; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

QY 12 GAAGACCCCTCAAGCGCAGCCGCTCAAAAAACCGACACCAAGTCATCACCAACCAAGCCAT 71  
 Db 95 GAGATCCCCAGGGAGATGCTGCCAGAGACAGATACATCCACCACCATGATCAGATCAC 154  
 QY 72 CGACTTTTAAATAAAATTAATCTCAAAATTTAGCCGAATTTGCTTTTCTTTTATAGACAA 131  
 Db 155 CCAACCTTCAACAAGATCACCCCAACCTGGCTGAGTTCCGCTTACGCTATATACGCCAG 214  
 QY 132 TTAGCTCATCAAAAGTAATTTCTACTACATTTTCTTTAGTCTGTTTCTATTGCCACTGCT 191  
 Db 215 CTGGCACACAGTCCCAACAGACCAATATCTTCTCTCCCAAGTACGATCGCTACAGCC 274  
 QY 192 TTCGCCATGTTGAGTTTAGTACTAAAGCCGATACCCATGAGGAGATTTTAAAGAGTTTA 251  
 Db 275 TTTGCAATGCTCTCCCTGGGACCAAGGCTGACACTCAGATGAATCTCTGGAGGCTG 334  
 QY 252 AACTTTAAATTTGACCGAAATCCAGAAAGCCCAAAATTCACGAGGGTTTTCACAGATTGTTG 311  
 Db 335 AATTTCAACCTCACGGAGATTCGGAGGCTCAGATCCATGAAGGCTTCCAGGAACCTCTC 394  
 QY 312 AGAATCTTGAATCAACCTGATTTCTCAATTTGCAATTAACCTACTGTTAAACGTTTATTTTG 371  
 Db 395 CGTACCTTCAACCAAGCCAGACAGCCAGCTCCAGCTGACCACCGGCANTGGCTGTTCTC 454  
 QY 372 TCTGAAGGTTTAAATTTGGTTGACAAATTTCTAGAAAGAGCTCAAGAAACATATATACAT 431  
 Db 455 AGCGAGGCTTGAAGCTAGTGGATAAGTTTGGAGGATGTTAAAAAGTTGTACCACTCA 514  
 QY 432 GAGGCTTTTACCGTTAAATTTTGGTGATCTAGGAAAGCTTAAAGCAAAATTAATGATTAT 491  
 Db 515 GAAGCCTTTCATCTCAACTTCGGGGACACCGAAGAGGCCCAAGAAACAGATCAACGATTAC 574  
 QY 492 GTTGAGAAAGGCCACCCAGGTAAGATCGTTGACCTAGTTTAAAGAAATAGATCGTGATACC 551  
 Db 575 GTGGAAGAGGTACTCAAGGGAATTTGGATTTGGTCAAGGAGCTTGACAGAGACACA 634  
 QY 552 GTCTTCGCACCTAGTTAACTATATTTTTCAGGGGTAAGTGGGAACGTCCTTTTCGAGGTT 611  
 Db 635 GTTTTGTCTCTGTGAATTAATACATCTTCTTTAAAGGCAATTTGGAGAGACCTTTGAAGTC 694  
 QY 612 AAAGATCACTCAAGAGNAGATTTTCATGTTGATCAAGTTACTACTCTCAAGTTTCCAAGT 671  
 Db 695 AAGGACACCGAGGAAGAGGACTTCCACGTGGACAGGTGACCCACCGGAGGTCCTATG 754  
 QY 672 ATGAAAGACTGGGTATGTTCAATATTAACATTTGCAAAATTAAGTTCTTTGGGTCTTA 731  
 Db 755 ATGGAGCGTTTAGCATGTTTAACATCCAGCACTGTGAAGAGCTGTCCAGCTGGGTGCTG 814  
 QY 732 TTAATGAAGTATTTAGGTAAAGCTACTGCTATTTTCTTTTACCACCAAGAGTAAAGCTT 791  
 Db 815 CTGATGAATACCTGGGGCATGCCACCGCATCTTCTTCTGCTGGATGATGAGGAAACTA 874  
 QY 792 CAACATTTAGAGAA 805  
 Db 875 CAGCACTGGGAAA 888



B0646142 982 bp mRNA linear EST 15-JUL-2002  
 LOCUS AGENCOURT\_8492569 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6296341  
 DEFINITION 5' mRNA sequence.  
 ACCESSION B0646142  
 VERSION B0646142.1 GI:21770314  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 982)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 Tissue procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM2504 row: a column: 14  
 High quality sequence stop: 647.  
 Location/Qualifiers  
 1..982  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NIH\_MGC\_100"  
 /issue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."  
 BASE COUNT 267 a 259 c 238 g 217 t 1 others  
 ORIGIN  
 Query Match 18.7%; Score 285.2; DB 14; Length 982;  
 Best Local Similarity 59.8%; Pred. No. 1.8e-63;  
 Matches 496; Conservative 0; Mismatches 333; Indels 1; Gaps 1;  
 QY 303 GAGTTGTTGAGAACTTGAATCAACCTGATTTCTCAATTGCAATTAAGTCTGTTAAGCGT 362  
 DB 1 GAACCTCTCCGTACCTCAACCCAGCAGACAGCCAGCTCCAGCTGACCCGGCAATGGC 60  
 QY 363 TTATTTTGTCTGAAGGTTTAAATTTGTTGACAAATTCCTAGAACGCTCAAGAACTA 422  
 DB 61 CTGTTCTCACGAGGGCCCTGAAGCTAGTGGATAAGTTTGTGGAGGATGTTAAAGATTG 120  
 QY 423 TATCATAGTAGGCTTTTACCGTTTAAATTTGGTGATCTAGTGAAGGAGCTAAAGCAAT 482  
 DB 121 TACCACCTAGAAGCTTCACTGTCTCACTCGGGACACCGAAGGAGCCAGAACAGATC 180  
 QY 483 AATGATTATGTTGAAGAAGGCCACCGGTAAGATCGTTGACCTAGTTAAAGAAATTAGAT 542  
 DB 181 AACGATTACGTGGAGAAAGGGTACTCAAGGGAATAATTTGGATTGGTCAAGGAGCTTGAC 240  
 QY 543 CGTGATACCGCTTTCGCACTAGTTAACTATATTTTTTCAAGGTAAGTGGGACGTCCT 602  
 DB 241 AGACACACAGTTTGTCTGGTGAATTACATCTCTTTAAAGGCAATATGGGAGAGACCC 300  
 QY 603 TTCGAGGTTAAAGTACTGAAGAGGAGATTTTCATGTTTATCATCAAGTTACTACTGTCAA 662  
 DB 301 TTTCAAGTCAAGGACACCGGAGGAGAGGACTTCCACGTGGACCCAGGTCACCCCGTGAAG 360

QY 663 GTTCCAATGATGAAGAAAGACTGGGTATGTTCATATTAACACATTGCAAAAAAATTAAGTTCT 722  
 DB 361 GTGCTATGATGAAGCGGTTTAGGCATGTTTAAACATCCAGCACTGTAAGAAGCTGTCAGC 420  
 QY 723 TGGGTCTTATTAATGAAGTATTTAGGTAAACGCTACTGCTATTTTTTTTACACAGCAA 782  
 DB 421 TGGGTGCTGCTGATGAATACCTGGCAATGCCACCCCATCTTCTTCTCCCTGATGAG 480  
 QY 783 GGTAAAGTTCAACATTTAGAGAATGAGTTGACTCATGACATTAATTAATTTTATAG 842  
 DB 481 GGGAACTACAGCACTTGGAAATGAACCTACCCACCATATCATCAACCAAGTTCTCTGAA 540  
 QY 843 AACGAGGATCGCTAGCGCTTCTGCACTGCCAAAGTTAAGTATCACCGGTACTTAC 902  
 DB 541 AATGAAGACAGAGGCTCTGCCAGCTTACATTTACCAAACTGTCCATTTACTTGAACCTAT 600  
 QY 903 GACTTAAATCTGTTTTAGGCCAGTTAGGTATTAACCAAGT-TTTTTCTAACGGTCCGA 961  
 DB 601 GATCTGAAGCGCTCTGGGTCACTGGGATCACTAAGGCTCTTTCAGCAATGGGCTGA 660  
 QY 962 TTTGAGTGGTGTACTGAAGAAGCTCCATTAATAATTGAGTAAAGCTGTTTCAAAAGCCGT 1021  
 DB 661 CCTCTCGGGTCAACAGAGGAGCACCCCTGAAGCTCTCCAGGCGCTGATTAAGGCTGT 720  
 QY 1022 CTTAACTATTGATGAAGAGGTACCGAGGCGCGGGCGGTATGTTCTGGAAGCTATTC 1081  
 DB 721 GCTGACCATCGACGAGAAAGGAGCTGAAGCTGCTGGGCCATGTTTGTAGAGCCATACC 780  
 QY 1082 AATGAGCATTCACACAGAAAGTTAAATTTAATAAACCATTCGTTTTTCTGA 1131  
 DB 781 CATGCTCTATCCCCCGGAGGCGAGTTCACCAAAACCCCTTGTGTTCTTTA 830

RESULT 9  
 LOCUS BM924019  
 DEFINITION AGENCOURT\_6709923 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5760183  
 5' mRNA sequence.  
 ACCESSION BM924019  
 VERSION BM924019.1 GI:19374398  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 985)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM12806 row: m column: 16  
 High quality sequence stop: 707.  
 Location/Qualifiers  
 1..985  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5760183"  
 /clone\_lib="NIH\_MGC\_116"  
 /lab\_host="DH10B"  
 /note="organ: pooled colon, kidney, stomach; Vector:  
 pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
 source anonymous pool of 3 colons, age 26 yo male, 49 yo  
 female, 71 yo male colon; 46 yo male kidney, and pool of 2  
 stomachs, 62 yo male and 70 yo female. Library is

FEATURES  
 source



Db 362 GTCAACTTCGGGACACCGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAGGCT 421  
 QY 504 ACCAGGGTAAGATCGTTGACCTAGTTAAAGAAATAGATCGTGATACCGCTTCGCGACTA 563  
 Db 422 ACTCAAGGGAAATGTGGATTGGTCAAGGAGCTTTGACAGAGACACAGTTTTTGTCTGTG 481  
 QY 564 GTTAACTATATTTTTTCAAGGTAAGTGGGAACGTCCTTTCAGGTTTAAAGATACCTGAA 623  
 Db 482 GTCAATTACATCTCTTTAAGGCAANTGGGAGAGACCCCTTTGAAGTCAAGGACACCCGAG 541  
 QY 624 GAGGAGATTTTCATGTTGATCAAGTTACTGTCTCAAAAGTTCCAAATGATGAAAGACTG 683  
 Db 542 GAAGAGGACTTCCACGTGGACCGAGGACCGCTGAAGTGCTTATGATGAAGCGTTTA 601  
 QY 684 GGTATGCTCAATATTCACATTCGCAAAAATTAAGTCTCTGGGTCTTATTAATGAAGTAT 743  
 Db 602 GGATGTTTAAACATCCAGCAGCTGTAAAGAGCTGTCCAGCTGGGTGCTGTGATGAATAC 661  
 QY 744 TTAGGTAAGGCTACTCTATTTTTTTTACCAGAGCAAGGTAAGCTTCCACATTTAGAG 803  
 Db 662 CTGGCAATGCCACCCCATCTCTCTGCTGATGAGGGGAACTACAGCACCTGGAA 721  
 QY 804 AATGACTTGACTCATGACATTAATTAATAATT--TTTAGAGAACGAGGATCGTCGTAGCG 861  
 Db 722 AATGAACCTCACCCAGCATATCATCACNAGTTCCTGGAAATGAAGACAGAAAGTCTGTG 781  
 QY 862 CTCTCTGACCTGCCAAAGTTAAGTATCACCGGTACTTACGACTTAAATCTGTTTAG 921  
 Db 782 CCACCTTAGATTTACCCAAACTGTCCATTACTGGAACACTATGATCTGAAGAGCGCTCGT 841  
 QY 922 GCAGGTTAGGTATACCAAAATTTTTTCTAACGGGTGCC 959  
 Db 842 GTCACTGGGCATCACTAAGGCTTCAGCAATGGGGC 879

## RESULT 11

BM924813  
 LOCUS AGENCOURT\_6653921 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5761267  
 DEFINITION 5' mRNA sequence.  
 ACCESSION BM924813  
 VERSION BM924813.1 GI:19375192  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 DNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM12809 row: j column: 20  
 High quality sequence stop: 662.  
 Location/Qualifiers  
 1..1194  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5761267"  
 /clone\_lib="NIH\_MGC\_116"  
 /lab\_host="DH10B"  
 /note="Organ: pooled colon, kidney, stomach; Vector: PCMV-SPORT0; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2

## FEATURES

source  
 1..1194  
 Db 727 GAAAAAG 732

## RESULT 12

BQ650189  
 LOCUS AGENCOURT\_8298326 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6269613  
 DEFINITION 5' mRNA sequence.  
 ACCESSION BQ650189  
 BQ650189 907 bp mRNA linear EST 15-JUL-2002

stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by c. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

BASE COUNT 321 a 303 c 328 g 242 t  
 ORIGIN

Query Match 18.0%; Score 274.8; DB 14; Length 1194;  
 Best Local Similarity 61.2%; Pred. No. 9.7e-61;  
 Matches 444; Conservative 0; Mismatches 282; Indels 0; Gaps 0;  
 QY 448 ATTTTGGTACTGAGGAAGCTAAAAGCAAATTAATGATTTGTTGAGAAAGCGACCC 507  
 Db 7 AATTTCCGGGATCCGAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAGGGTACTC 66  
 QY 508 AGGGTAAGATCGTTGACCTAGTTAAAGNATTTAGATCGTGATACCGTCTTCCACATAGTTA 567  
 Db 67 AAGGAAAATTTGTGGATTGTGTCAGGAGCTTTGACAGAGACACAGTTTTTGTCTGTGTA 126  
 QY 568 ACTATATTTTTTCAAGGGTAAGTGGGAACGTCCTTTTCGAGGTTTAAAGATACTCAAGAGG 627  
 Db 127 ATTACATCTCTTTAAAGGCAATTTGGGAGAGACCCCTTTGAAGTCAAGACACACCGAGGA 186  
 QY 628 AAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTTTCCAAATGATGAAAGAGCTGGGTA 687  
 Db 187 AGGACTTCCACGTGGACAGGTGACCACCGTGAAGGTGCTTATGATGAAGCGTTTAGGCA 246  
 QY 688 TGTTCATATTTCAACATTTGCAAAAATTAAGTCTTGGGTCTTATTAATGAAGTATTTAG 747  
 Db 247 TGTTTACATCCAGCACTGTAAAGAAAGCTGTCCAGCTGGGTGCTGCTGATGAATACCTGG 306  
 QY 748 GTAACGCTACTGCTATTTTTTTTTTACCAGACGAAGGTAAGCTTCAACATTTAGAGAATG 807  
 Db 307 GCATGCGACCGCCATCTTCTCTGCTGATGAGGGGAAACTACAGCACTGGAAATG 366  
 QY 808 AGTTGACTCATGACATTAATTAATAATTTTTPAGAGAACGAGGATCGTGTAGCGCTTCTC 867  
 Db 367 AACTCACCCAGCATATCATCAACCAAGTTCTCTGGAAATGAAGACAGAAGGTCTGCCAGCT 426  
 QY 868 TGCACCTGCCAAAGCTTAAGTATCACCGTACTTACGACTTAAATCTGTTTAGGCCAGT 927  
 Db 427 TACATTTACCCAAACTGTCCATTTACTGGAACCTATGATCTGAAGAGCGCTCTGGGTCAAC 486  
 QY 928 TAGTATTAACAAAGTTTTTTTCTAACGGTGGCGGATTTGAGTGGTGTGTTACTGAAGAAGCTC 987  
 Db 487 TGGGCATCACTAAGCTCTTCAGCAATGGGCTGACCTCTCCGGGGTCAAGAGAGGACAC 546  
 QY 988 CATTAAATTTGAGTAAAGCTGTTTCCAAAGCCGCTTTAACTATTTGATGAAAGGGGTACCG 1047  
 Db 547 CCCTGAAGCTCTCCAAGGCGGTGCATTAAGGCTGTGCTGACCATCGACGAGAGAGGACTG 606  
 QY 1048 AGGCGCGCGGCGCTATCTTCTGGAAGCTATTCCAATGAGCATTTCCACCAGAGTTAAAT 1107  
 Db 607 AAGTGTCTGGGGCCATGTTTTTAGAGGCCATACCCATTTGTCATCCCCCGGAGGTCAAGT 666  
 QY 1108 TTAATAAACCATTCGTTTTTTTCTGATGATCGAGAGAACACATTAAGAGCCCATTTGTTATGG 1167  
 Db 667 TCAACAAACCTTTGTCTTCTTAATGATTGAACAANAATACCAAGTCTCCCCCTCTTCATGG 726  
 QY 1168 GTAAGG 1173  
 Db 727 GAAAAAG 732

Mon Dec 9 12:50:36 2002

```
VERSION B0650189.1 GI:21774361
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 907)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov
          Tissue Procurement: CGAP (Stanford)
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM2443 row: 9 column: 22
          High quality sequence stop: 650.
FEATURES
    source
    1..907
    Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:6269613"
        /clone_lib="NIH_MGC_100"
        /tissue_type="hepatocellular carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: liver; Vector: pOTB7; Site:1; XhoI; Site:2;
          EcoRI; cDNA made by oligo-dT priming. Directionally cloned
          into EcoRI/XhoI sites using the following 5' adaptor:
          GGACACGAG(G). Size-selected >500bp for average insert size
          1.8kb. Library constructed by Ling Hong in the laboratory
          of Gerald M. Rubin (University of California, Berkeley)
          using ZAP-cDNA synthesis kit (Stratagene) and Superscript
          II RT (Life technologies). Note: this is a NIH_MGC
          Library."
BASE COUNT 245 a 237 c 226 g 198 t 1 others
ORIGIN
Query Match 17.9%; Score 273.4; DB 14; Length 907;
Best Local Similarity 59.4%; Pred. No. 2.le-60;
Matches 498; Conservative 0; Mismatches 337; Indels 3; Gaps 2;
QY 336 CAATTGCAATTAACTACTGTTAAACGGTTATTTTGTCTGAGGTTTAAAAATGTTGAC 395
DB 7 GAGCTCCAGCTGACACCGCAATGCGCTGTTCTCAGCGAGGGCCTGAAGCTAGTGGAT 66
QY 396 AAATTCCTAGAGAGCTCAAGAACTATATCATAGTAGGCTTTTACCGTTAAATTTGGT 455
DB 67 AAGTTTTTGGAGGATGTTAAAAAGTTGTACCACCTCAGAAAGCCTTCACTGTCACACTTCGGG 126
QY 456 GATAGTGAAGAGCTAAAGCAATTAATGATTATTTTGAGAAAGGACCCAGGTTAAG 515
DB 127 GACACCGAAGGCCAAGAAACAGATACACGATTACGTTGAGAGAGGTTACTACAGGGAAA 186
QY 516 ATCGTTGACCTAGTTAAAGAAATAGATCGTGATACCGTTCTGCGACTAGTTAACTATAT 575
DB 187 ATGTGGAATTTGGTCAAGAGCTTGACAGAGACACAGTTTTTGTCTGTTGAATATACATC 246
QY 576 TTTTCAAGGGTAAGTGGGAAGCTGCTTTCGAGGTTAAAGATACATGAAGAGGAATTTT 635
DB 247 TTTCTTTAAAGGCAATGGGAGAGACCTTTTGAAGTCAAGGACACCGAGAGAGGACATTC 306
QY 636 CATGTTGATCAAGTTACTACTGTCAAAGTTCCAAATGTTAAAGAAAGTGGTATGTTCAAT 695
DB 307 CAGGTGGACAGGTGACACCGCTGAAGGTCCTATGATGAACCGTTAGGCAATTTTAC 366
QY 696 ATTCACATTTGCAAAAAATTAAGTTCTTGGGCTTTATTAATGAAGTATTTAGGTAAACGT 755
DB 367 ATCCAGCACTGTAAAGAGCTGTCACAGCTGGTGTCTGATGAATATACCTGGGCAATGCC 426
```

```
BASE COUNT      246 a      276 c      245 g      192 t
ORIGIN
Query Match      17.7%; Score 270.4; DB 14; Length 959;
Best Local Similarity 59.3%; Pred. No. 1.3e-59;
Matches 477; Conservative 0; Mismatches 326; Indels 1; Gaps 1;
QY 186 ACTGCTTCGCCGATGTTAGTTAGTACTAAAGCGGATACCCTAGCAGAGATTTTAAAG 245
    || || || || || || || || || || || || || || || || || || || || ||
Db 1 ACAGCCTTTTGCAATGCTCTCCTGGGGACCAAGGCTGACACTCAGATGAATCTGGAG 60
QY 246 GGTTTAACTTTAAATTTGACCAAAATCCAGAAAGCCCAATTCACGAGGGTTTTCAGAG 305
    || || || || || || || || || || || || || || || || || || || || ||
Db 61 GGCCTCAATTTCAACCTTCACGAGATTCGGAGGCTCAGATCATCAAGGCTTCAGGAA 120
QY 306 TTGTTGAGAACTTTGAATCAACCTGATCTCAATTTGCAATTAACACTGCTGTAACGGTTTA 365
    || || || || || || || || || || || || || || || || || || || || ||
Db 121 CTCCTCCGTAACCTCAACGAGCAGACAGCCAGCTCCAGCTCACCACCGCAATGGCCTG 180
QY 366 TTTTCTGCTGAAGGTTTAAATTTGGTTGACAAATTCCTAGAAGAGCTCAAGAAACTATAT 425
    || || || || || || || || || || || || || || || || || || || || ||
Db 181 TTCTCAGGAGGGCTGAAGCTAGTGGAATAAGTTTTTGGAGGATGTTTAAAGTTGTAC 240
QY 426 CATGATGAGGCTTTTACCCTTAATTTTGGTGATCTAGGAAAGCTAAAGCAAAATTAAT 485
    || || || || || || || || || || || || || || || || || || || || ||
Db 241 CACTCAGAAGCCTTCACTGTCAACTTCGGGGACACCGGAAGAGGCCAAGAAACAGATCAAC 300
QY 486 GATTATGTTGAGAAAGGACCCAGGTAAGATCGTTGACCTAGTTAAAGAAATTAAGATCGT 545
    || || || || || || || || || || || || || || || || || || || || ||
Db 301 GATTACGTGGAGAGGGTACTCAAGGAAATTTGGATTTGGTCAAGGAGCTTGACAGA 360
QY 546 GATACCGTCTTCGCACTAGTTAACTATATTTTTTCAAGGGTAAGTGGGAAGCTCCTTTC 605
    || || || || || || || || || || || || || || || || || || || || ||
Db 361 GACACAGTTTTTGCCTGTGTGAATTAACATCTCTTTAAAGGCAATGGGAGAGACCCCTT 420
QY 606 GAGGTTAAAGATCTGAAGAGGAAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTT 665
    || || || || || || || || || || || || || || || || || || || || ||
Db 421 GAAGTCAAGGACACCGGAGGAGGAGGACTTCCACGTGGACCAAGTGCACCACCGTGAAGGTG 480
QY 666 CCAATGATCAAAAGACTGGGTATGTTCAATATCAACATTTGCAAAATTAAGTTCTTGG 725
    || || || || || || || || || || || || || || || || || || || || ||
Db 481 CCTATGATGAAGGTTTGGCTATTTTAACTCCAGCACTGTGAAGAGCTGTCAGCTGG 540
QY 726 GTCTTATTAATGAAGTATTAGGTAAAGCTACTGCTATTTTTTTTACCAGACGAAGT 785
    || || || || || || || || || || || || || || || || || || || || ||
Db 541 GTGCTCTGATGAATACCTTGGCAATGCCACGCCATCTCTCTCTGCTGATGAGGG 600
QY 786 AGCTTCAACATTTAGAGAAATGAGTTGACTCATGACATTAATTAATAATTTTGAAGAAC 845
    || || || || || || || || || || || || || || || || || || || || ||
Db 601 AAACATACAGCAGCTGGAAATGAACCTACCCACGATATCATCAACCAAGTTCTCTGGAAAT 660
QY 846 GAGGATCGTGTAGCGCTTCTCTGCACCTGCCAAAGTTAAGTATCACCGGTACTTACGA- 904
    || || || || || || || || || || || || || || || || || || || || ||
Db 661 GAAGACAGAGGTCGCCAGCTTACATTTACCACCAACTGTCCTATTGGAACCTATGAT 720
QY 905 CTTAAATCTGTTTATAGCCAGTTAGGTATTACCAAGTTTCTTACAGGGTGCAGATT 964
    || || || || || || || || || || || || || || || || || || || || ||
Db 721 CTGAACAGCGCTCTGGGTCAACTGGGCATCACTAAGGCTTCAGCCATGGGGCTGGACT 780
QY 965 GAGTGTGTTTACTGAAGAGCTCC 988
    || || || || || || || || || || || || || || || || || || || || ||
Db 781 CCCCCGGGCTCACAGAGGAACGCC 804
```

```
RESULT 14
BQ646948
LOCUS      BQ646948      891 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8302495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271313
            5', mRNA sequence.
ACCESSION  BQ646948
VERSION    BQ646948.1
KEYWORDS   EST.
SOURCE     human.
```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM2447 row: n column: 18  
High quality sequence stop: 672.

FEATURES  
source  
1..891  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6271313"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOH7; Site\_1: XhoI; Site\_2:  
ECORI; cDNA made by oligo-dT priming. Directionally cloned  
into EORI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT 229 a 247 c 226 g 189 t

Query Match 17.68; Score 268.6; DB 14; Length 891;  
Best Local Similarity 60.3%; Pred. No. 3.7e-59;  
Matches 479; Conservative 0; Mismatches 314; Indels 2; Gaps 2;

QY 12 GAAGACCTCTAAGCGGACGGCTCAAAAACCCAGACCATCATCAGCACCAAGACCAT 71  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 88 GAGGATCCCCAGGAGATGCTGCCAGAGACAGATACATCCCACCATGATCAGATCAC 147  
QY 72 CCGACTTTTAAATAAATTTACTCCAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA 131  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 148 CCAACCTTCAACAAGATCACCCCAACCTGGCTGAGTTCGCCCTTACGCCCTATACGCCAG 207  
QY 132 TTAGCTCATCAAACTAATTTCTACTAATTTTCTTTTAGTCTGCTTTCTATTGCCACTGCT 191  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 208 CTGGCACACCATGCTCCACAGACCAATATCTTCTTCCCAGTGAGCATCGCTACAGCC 267  
QY 192 TTCGCCATGTTGAGTTTATAGTACTAAAGCCGATACCCATGACGAGATTTTAGAAGGTTTA 251  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 268 TTTGCAATGCTCTCCCTGGGACCAAGCTGACACTCAGCATGAATCTCTGGAGGGCTG 327  
QY 252 AACTTTAAATTTGACCGAAATCCCGAGAGCCCAATTTACGAGGGTTTTCAGAGTTGTTG 311  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 328 AATTTCAACCTCACGGAGATTCGGGAGGCTCAGATCATGAAGGCTTCCAGGAACCTCCTC 387  
QY 312 AGAATTTGAATCAACCTGATTTCTCAATTTGAATTAATCTACTGTAACGGTTTATTTTGG 371  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 388 CGTACCTCAACAGCAGCAGACAGACGCTCCAGCTGACCCGCAATGGCTGTTCTCTC 447  
QY 372 TCTGAAGTTTAAATTTGGTTGACAAATTTCTAGAGACGCTCAAGAACTATATCATPATG 431  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 448 AGCGAGGGCTTGAAGCTAGTGAGTAAGTTTGTGGAGGATGTTTAAAGTTGTACCACTCA 507  
QY 432 GAGGCTTTTACCGTTAATTTTGGTGATCTAGGAGAACCTAAAGCAAAATTAATGATTAT 491  
 || || || || || || || || || || || || || || || || || || || || ||  
Db 508 GAAGCCTTCACTGTCAACTTCGGGGACACCCGAGAGGCCAAGAACAGATCAACGATTAC 567

